

CLAIMS

What is claimed is:

1. A morphological analyzer for performing a morphological analysis on a natural language text to be processed, comprising:

a dictionary unit storing header words and attribute information of the header words;

a token list generating unit for referencing data in said dictionary unit, extracting tokens that can form the natural language text from said natural language text to be processed, and registering them on a token list; and

a token string selecting unit for selecting optimum token strings for composing said natural language text on the basis of the token list generated by said token list generating unit,

wherein said token list generating unit controls the registration of the tokens on said token list on the basis of conditions imposed on the morphological analysis and said attribute information of the header words corresponding to said extracted tokens.

2. The morphological analyzer according to Claim 1, wherein said token list generating unit registers, on said token list, only the tokens having attributes matching said conditions imposed on the morphological analysis on the basis of the attribute information of said header words corresponding to said tokens.

3. The morphological analyzer according to Claim 1, wherein said dictionary unit stores the information indicating whether or not said header words are decomposable as the attribute information of said header words, said token list generating unit references the attribute information of said header words, and registers the extracted tokens on said token list, except tokens corresponding to decomposable header words from said extracted tokens, when a condition of decomposing the complex word for making the morphological analysis is imposed.

4. The morphological analyzer according to Claim 1, wherein the attribute information of said header words stored in said dictionary unit is recorded with the flag data having a number of bits corresponding to the number of attribute information, and said token list generating unit decides whether or not to register said tokens on said token list on the basis of the value of said flag data for said header words corresponding to said tokens.

5. A morphological analyzer for performing a morphological analysis on a natural language text to be processed, comprising:

token list generation means for decomposing said natural language text to be processed into tokens that are components of the natural language text and registering them on a token list except tokens decomposable into smaller tokens; and

token string selection means for selecting optimum token strings for composing said natural language text on the basis of the token list generated by said token list generation means.

6. The morphological analyzer according to Claim 5, wherein said token list generation means selectively controls whether or not tokens decomposable into smaller tokens are excluded from tokens registered on said token list in accordance with the given conditions imposed on the morphological analysis.

7. A natural language processor, comprising:

morphological analysis means for performing a morphological analysis on a natural language text to be processed; and

application execution means for performing given processing for said natural language text morphologically analyzed by said morphological analysis means, said morphological analysis means comprising:

a dictionary unit storing header words and attribute information of the

header words;

a token list generating unit for referencing data in said dictionary unit, extracting tokens that can form the natural language text from said natural language text to be processed, and registering only tokens having attributes matching conditions requested by said application execution means on a token list on the basis of the attribute information of said header words corresponding to the tokens; and

a token string selecting unit for selecting optimum token strings for composing said natural language text on the basis of the token list generated by said token list generating unit.

8. The natural language processor according to Claim 7, wherein said dictionary unit stores the information indicating whether or not said header words are decomposable as the attribute information of said header words, said token list generating unit references the attribute information of said header words, and registers tokens corresponding to undecomposable header words on said token list, when it is requested by said application execution means to decompose decomposable words for making the morphological analysis.

9. The natural language processor according to Claim 7, wherein the attribute information of said header words stored in said dictionary unit is recorded with the flag data having a number of bits corresponding to the number of attribute information, and said token list generating unit decides whether or not to register said tokens on said token list on the basis of the value of said flag data for said header words corresponding to said tokens.

10. A morphological analysis method of performing a morphological analysis on a natural language text by using a computer, comprising the steps of:

inputting a natural language text to be processed, referencing a dictionary stored in a memory, obtaining tokens that can form the natural language text and attribute

information of the tokens, and storing them in a work area of the memory;

selecting given tokens out of the tokens stored in said memory on the basis of given conditions imposed on the morphological analysis and said attribute information of the tokens and registering them on a token list formed in a given area of the memory;

generating token strings that can form said natural language text to be processed on the basis of said token list and storing them in the work area of the memory; and

selecting optimum token strings for composing said natural language text to be processed out of said token strings stored in said memory and outputting them.

11. The morphological analysis method according to Claim 10, wherein a step of registering said tokens on said token list includes registering only tokens having attributes matching the given conditions on said token list in accordance with said given conditions imposed on said morphological analysis.

12. A morphological analysis method of performing a morphological analysis on a natural language text by using a computer, comprising the steps of:

inputting a natural language text to be processed, decomposing it into tokens that are components of the natural language text, and storing the obtained token group in a work area of a memory;

registering said token group on a token list formed in a given area of the memory except tokens decomposable into smaller tokens;

generating token strings that can form said natural language text to be processed on the basis of said token list and storing them in the work area of the memory; and

selecting optimum token strings for composing said natural language text to be processed out of said token strings stored in said memory and outputting them.

13. A program for controlling a computer to perform a morphological analysis on a natural language text, the program enabling said computer to serve as:

means for referencing a dictionary having records of header words and attribute

information of the header words and stored in a given storage device, extracting tokens that can form the natural language text from said natural language text to be processed, selecting given tokens from the extracted tokens on the basis of given conditions imposed on the morphological analysis and attribute information of said tokens, and registering them on a token list formed in a given area of a memory; and

means for selecting optimum token strings for composing said natural language text on the basis of the token list generated by a token list generating unit.

14. The program according to Claim 13, wherein means for registering said tokens on said token list decides whether or not to register said tokens on said token list on the basis of the value of flag data having a number of bits corresponding to the number of attribute information of said tokens recorded in said dictionary.

15. A program for controlling a computer to perform a morphological analysis on a natural language text, the program causing said computer to execute:

a first process of inputting a natural language text to be processed, decomposing it into tokens that are components of the natural language text, and storing the obtained token group in a work area of a memory;

a second process of registering said token group on a token list formed in a given area of the memory except tokens decomposable into smaller tokens;

a third process of generating token strings that can form said natural language text to be processed on the basis of said token list and storing them in the work area of the memory; and

a fourth process of selecting optimum token strings for composing said natural language text to be processed out of said token strings stored in said memory and outputting them.

16. The program according to Claim 15, wherein said program causes said computer to execute a process of judging the given conditions imposed on the morphological

analysis, and a process of registering all said tokens on said token list in accordance with said given conditions, instead of said second process.